A STRONG LAW FOR SOME GENERALIZED URN PROCESSES

by

Bruce Hill, David Lane, and William Sudderth*

Technical Report No. 326

September, 1978

* Research supported by National Science Foundation Grant MCS77-28424.

Abstract

Let f be a continuous function from the unit interval to itself and let $X_0, X_1, \ldots$ be the successive proportions of red balls in an urn to which at the n^{th} stage a red ball is added with probability $f(X_n)$ and a black ball with probability $1-f(X_n)$. Then X_n converges almost surely to a random variable X with support contained in the set $C = \{ p: f(p)=p \}$. If, in addition, $0 < f(p) < 1$ for all p, then, for each r in C, $P[X=r] > 0$ $(= 0)$ when $f'(r) < 1$ (> 1). These results are extended to more general functions f.

1. Introduction

Let $X_0 = x$ be the proportion of red balls in an urn containing m balls altogether and let f be a mapping from the unit interval into itself. Suppose that a red ball is added to the urn with probability $f(x)$ and a black ball is added with probability $1-f(x)$. Let X_1 be the new proportion of red balls and iterate the procedure to generate an <u>urn process</u> X_1, X_2, ... This paper is concerned with two questions about the asymptotic behavior of such urn processes. Does X_n converge almost surely? And, if so, what is the support of the limit variable?

The distribution of $\{X_n\}$ is determined by the <u>urn function</u> f and the <u>initial urn composition</u> (x,m). The process $\{X_n\}$ is of course Markov, with nonstationary transition probabilities. Only the values of f on the rationals play a role in the transitions of $\{X_n\}$ but it is convenient to think of f as defined on the whole unit interval.

Here are three examples of urn processes whose asymptotic behavior is already well understood.

<u>Example 1.1</u> Bernoulli urns.

Let $0 \leq p_o \leq 1$ and suppose $f(p) = p_o$ for all p in $[0,1]$. Then

$$X_n = (mx + S_n)(m + n)^{-1}$$

where $S_n = Y_1 + \ldots + Y_n$ for each n and $Y_1, Y_2, \ldots$ is a sequence of independent variables each being Bernoulli (p_o); that is, each equalling 1 and 0 with probabilities p_o and $1-p_o$ respectively. By the strong law of large numbers, X_n converges to p_o almost surely.

<u>Example 1.2</u> A Bernard Friedman urn.

If $f(p) = 1-p$, then X_n converges to 1/2 almost surely (Freedman, 1965).

In both examples 1.1 and 1.2, the limiting distribution is independent of the initial urn composition. This is not always true, as the next example shows.

Example 1.3 Polya urns.

If $f(p) = p$ for all p, then $\{X_n\}$ is a Polya process (Johnson and Kotz (1977), Chapter 4) and converges almost surely to a random variable X whose distribution is absolutely continuous with respect to Lebesgue measure. Indeed X has a Beta distribution with parameters $s = mx$ and $m - s$, and has the density

$$(1.1) \qquad \varphi_{s,m-s}(p) = c_{s,m-s} p^{s-1}(1-p)^{m-s-1},$$

where $c_{s,m-s} = (m-1)!/(s-1)!(m-s-1)!$
and $0 < p < 1$. (For a proof of convergence, see Freedman (1965).)

In all the examples, the urn process converges almost surely to a limit variable X . This is a quite general phenomenon: in particular, it happens whenever the set of discontinuities of the urn function f is nowhere dense in [0,1] (Corollary 2.1; Theorem 2.1 gives a more general result). However, there exist discontinuous urn functions whose associated urn processes almost surely do not converge (Example 2.1).

In the three examples above, the limit variable X has support equal to the crossing set $C = \{p: f(p) = p\}$. For continuous urn functions, the support of X is always contained in C (Corollary 3.1), but need not in general be equal to C.

Example 1.4

Suppose $0 < p_o < 1$, f is continuous, $f(p) < p$ for $0 < p < p_o$, and $f(p) > p$ for $p_o < p < 1$. Then $C = \{0, p_o, 1\}$, but p_o is not contained in the support of X as follows from Theorem 5.1.

The essential difference between p_o in the first and last examples is that between a downcrossing and an upcrossing by f of the diagonal. A point p_o in $[0,1]$ is an <u>upcrossing</u> (<u>downcrossing</u>) if, for all p in some neighborhood of p_o, $p < p_o$ implies $f(p) < p$ $(f(p) > p)$ and $p > p_o$ implies $f(p) > p$ $(f(p) < p)$. In particular, if f is differentiable at a point p_o in C, then p_o is an upcrossing (downcrossing) point if and only if $f'(p_o) > 1$ $(f'(p_o) < 1)$. In general, downcrossing points are limit points for urn processes (Theorem 4.2), but urn processes never converge to an upcrossing point (Theorem 5.1). Indeed, in the special case when the crossing set consists only of upcrossings and downcrossings and the urn function is continuous and does not assume the values 0 and 1, the limit distribution of the urn process has support precisely equal to the set of downcrossings (Theorem 6.1).

2. <u>A convergence theorem</u>.

Suppose f is an urn function. Define the <u>diagonal oscillation set</u> $\mathfrak{G} = \{ p$: in every neighborhood of p there exist points p_1, p_2 such that $f(p_1) < p_1$ and $f(p_2) > p_2\}$. The following result will be proved in this section.

<u>Theorem 2.1</u>: If $\mathfrak{G} \cap C^c$ is nowhere dense in $[0,1]$, then X_n converges almost surely.

Throughout the paper $\{X_n\}$ has urn function f and initial urn composition (x,m) unless explicitly stated otherwise.

The proof of the theorem will follow five lemmas. The first lemma is just a statement of the strong Markov property for urn processes. As usual, a stopping time T for $\{X_n\}$ is a random variable with values in $\{0, 1, \ldots\} \cup \{\infty\}$ such that, for every n, the event $[T \le n]$ is in the σ-field $\mathfrak{F}_n$ generated by $\{X_0, \ldots, X_n\}$. The σ-field $\mathfrak{F}_T$ consists of those events A such that $A \cap [T \le n] \in \mathfrak{F}_n$ for all n.

Lemma 2.1: If T is an almost surely finite stopping time for $\{X_n\}$, then the conditional distribution of $\{X_T, X_{T+1}, \ldots\}$ given $\mathfrak{F}_T$ is the distribution of an urn process which has initial composition $(X_T, m + T)$ and the same urn function f.

The Markov property will often be used without reference to Lemma 2.1.

The next lemma makes possible the pathwise comparison of two urn processes. For its statement, let $\{Y_n\}$ be an urn process with urn function g and initial composition (y,m).

Lemma 2.2: If $f \le g$ and $x \le y$, then $\{X_n\}$ and $\{Y_n\}$ can be realized on a common probability space in such a way that $X_n \le Y_n$ for all n.

Proof: Let $n \ge 0$ and suppose $\{X_0,\ldots,X_n\}$ and $\{Y_0,\ldots,Y_n\}$ have been constructed so that $X_i \le Y_i$ for $0 \le i \le n$.

To define (X_{n+1}, Y_{n+1}), introduce random variables (A_n, B_n) satisfying:

(i) A_n is Bernoulli $(f(X_n))$ and B_n is Bernoulli $(g(Y_n))$.

(ii) The pair (A_n, B_n) is conditionally indpendent of $(X_0,\ldots,X_{n-1}, Y_0,\ldots,Y_{n-1})$ given (X_n, Y_n).

(iii) If $X_n = Y_n$, then $A_n \le B_n$.

It is easy to verify that such variables exist. (For part (iii), use the fact that, if $X_n = Y_n$, then $f(X_n) = f(Y_n) \le g(Y_n)$.)

Now set

$$X_{n+1} = \frac{(m+n)X_n + A_n}{m + n + 1} ,$$

$$Y_{n+1} = \frac{(m+n)Y_n + B_n}{m + n + 1} .$$

Then $X_{n+1} \le Y_{n+1}$, and $\{X_0,\ldots,X_{n+1}\}$ and $\{Y_0,\ldots,Y_{n+1}\}$ have the desired joint distributions. □

The next lemma shows that, in the case when the initial urn size m is large, a Polya process is unlikely to venture far from its initial position. In the sequel, $P_{(x,m)}$ is sometimes used to denote the distribution of the urn process with initial urn composition (x,m) and fixed urn function f.

<u>Lemma 2.3</u>: Suppose $f(x) = x$ for all x. Then, for every $\epsilon > 0$ and initial composition (x,m),

$$P_{(x,m)}[\sup_n |X_n - x| \ge \epsilon] \le \frac{(s+1)(m-s+1)}{\epsilon^2(m+2)^2(m+1)}, \tag{2.1}$$

where $s = mx$.

<u>Proof</u>: As mentioned in Example 1.3, the martingale X_n converges almost surely to a random variable X which has a Beta $(s, m-s)$ distribution. The right-hand-side of (2.1) is just $E(X-x)^2/\epsilon^2$ and (2.1) is the familiar Kolmogorov inequality for the submartingale $\{(X_n-x)^2\}$. □

Let $I = (a,b)$ and $J = (c,d)$ where $a < c < d < b$. Let E be the event that $\{X_n\}$ upcrosses I infinitely often, that is, the event that $X_n \le a$ for infinitely many n and $X_n \ge b$ for infinitely many n.

Lemma 2.4: If $P_{(x,m)}(E) > 0$, then, for every $\varepsilon > 0$ and positive integer N, there exist $y \in J$ and $n \geq N$ such that $P_{(y,n)}(E) \geq 1 - \epsilon$.

Proof: By Levy's martingale convergence theorem,

$$P_{(x,m)}(E|X_1, \ldots, X_n) \to 1_E$$

almost surely. By the Markov property,

$$P_{(x,m)}(E|X_1 = x_1, \ldots, X_n = x_n) = P_{(x_n,m+n)}(E).$$

Now use the fact that any path in E must visit J infinitely often because the increments of $\{X_n\}$ go uniformly to zero. □

Lemma 2.5: If $f(p) = p$ for all $p \in I$, then

$$\lim_{n\to\infty} \sup_{y\in J} P_{(y,n)}[X_n \text{ visits } I^c] = 0, \tag{2.2}$$

and

$$P_{(x,m)}(E) = 0 \tag{2.3}$$

for every initial composition (x,m).

Proof: The first equality follows from Lemma 2.3. The second is a consequence of the first together with Lemma 2.4. □

Proof of Theorem 2.1: By a familiar argument, it suffices to show $P_{(x,m)}(E) = 0$. Suppose, to the contrary, that $P_{(x,m)}(E) > 0$. Consider two cases.

Case 1. There is a nondegenerate interval $I^1 \subseteq I$ such that $f(p) = p$ for all $p \in I^1$.

Let E^1 be the event that $\{X_n\}$ upcrosses I^1 infinitely often. Clearly, $E \subseteq E^1$. But, by Lemma 2.5, $P_{(x,m)}(E^1) = 0$.

Case 2. For every nondegenerate interval $I^1 \subseteq I$, there exists $p \in I^1$ such that $f(p) \neq p$.

There must exist a nondegenerate subinterval $I^1 \subseteq I$ such that either $f(p) < p$ for all $p \in I^1$ or $f(p) > p$ for all $p \in I^1$ because otherwise $\mathfrak{G}$ would be dense in the interval I. To be specific, suppose that $f(p) < p$ for all $p \in I^1$. (The other case is similar.) Define S to be the supremum of the X_n's. The event $[S \geq b]$ contains E so that, by Lemma 2.4,

$$\lim_{n\to\infty} \sup_{y \in J} P_{(y,n)}[S \geq b] = 1. \tag{2.4}$$

Next consider the urn function g given by

$$g(p) = f(p) \text{ if } p \notin I$$
$$= p \text{ if } p \in I.$$

Then $f \leq g$ and, by Lemma 2.2, the supremum of an f-process is stochastically smaller than that of a g-process with the same initial composition. Thus (2.4) holds also for g-processes. But, by Lemma 2.5, the contradictory equality (2.2) also holds for g-processes. □

Corollary 2.1. If the set of discontinuities of f is nowhere dense, then X_n converges almost surely.

Proof: Each point in $\mathfrak{G} \cap C^c$ is a discontinuity point of f. □

Here is an example of an urn process which does not converge.

Example 2.1. Define

$$S = \{(2^k + j)(2^{k+1} + j)^{-1}: k = 0,1,\ldots; \quad j = 0,1,\ldots, 2^k - 1\}$$

and

$$T = \{2^{k+1}(2^{k+1} + 2^k + j)^{-1}: k = 0,1,\ldots; \quad j = 0,1,\ldots, 2^k - 1\}.$$

It can be shown that S and T are disjoint so that there can be unambiguously defined an urn function f which equals 1 at each point of S and equals 0 at each point of T. The corresponding urn process with initial composition $(2^{-1}, 2^{k+1})$ moves deterministically as indicated below:

$$\frac{1}{2} = \frac{2^k}{2^{k+1}} \rightarrow \frac{2^k + 1}{2^{k+1} + 1} \rightarrow \cdots \rightarrow \frac{2^k + 2^k}{2^{k+1} + 2^k} = \frac{2}{3}$$

$$\frac{2}{3} = \frac{2^{k+1}}{2^{k+1} + 2^k} \rightarrow \frac{2^{k+1}}{2^{k+1} + 2^k + 1} \rightarrow \cdots \rightarrow \frac{2^{k+1}}{2^{k+2}} = \frac{1}{2}$$

Hence the process oscillates between 1/2 and 2/3 forever.

3. Convergence to the diagonal.

Suppose f is an arbitrary urn function, and the value of X_n is $y = s(m+n)^{-1}$. Then the conditional expectation of X_{n+1} given $[X_n = y]$ is calculated as follows:

$$(3.1) \qquad E(X_{n+1} \mid X_n = y) = (m+n+1)^{-1}[f(y)(s+1) + (1-f(y))s]$$
$$= y + (m+n+1)^{-1} (f(y)-y).$$

That is, the conditional mean of the increment $X_{n+1} - X_n$ is positive, zero, or negative depending on whether the graph of the urn function at X_n is above, on, or below the diagonal. Proposition 3.1 establishes an important consequence of this fact. To state the proposition, define, for every $\epsilon > 0$, the sets

$$A\epsilon = \{ x: \; f(x) - x > \epsilon \}$$

$$B\epsilon = \{ x: \; f(x) - x < \epsilon \}.$$

Proposition 3.1:

a) If $X_0 \in A\epsilon$, then $P[\{X_n\}$ exits from $A\epsilon] = 1$.

b) If $X_0 \in B\epsilon$, then $P[\{X_n\}$ exits from $B\epsilon] = 1$.

Proof: To prove (a), let T be the time of first exit of $\{X_n\}$ from $A\epsilon$ (or ∞ if the process never leaves $A\epsilon$). Let T_n be the minimum of T and n. T and T_n are stopping times for $\{X_n\}$. In particular, the event $[T_n \geq k]$ is $\mathfrak{F}_{k-1}$ - measurable. Thus

$$(3.2) \qquad 1 \geq E(X_{T_n}) \geq E(\sum_{k=1}^{n} (X_k - X_{k-1}) 1_{[T_n \geq k]})$$

$$= E(\sum_{k=1}^{n} E(X_k - X_{k-1} | \mathfrak{F}_{k-1}) 1_{[T_n \geq k]})$$

$$\geq E(\sum_{k=1}^{n} E(X_k - X_{k-1} | \mathfrak{F}_{k-1}) 1_{[T = \infty]}).$$

Using Lemma 2.1 and (3.1),

$$(3.3) \qquad E(X_k - X_{k-1} | \mathfrak{F}_{k-1}) = E(X_k - X_{k-1} | X_{k-1})$$

$$\geq \epsilon \; [m+k]^{-1} ,$$

for X_{k-1} in $A\epsilon$.

Combining (3.2) and (3.3), for every n,

$$(3.4) \qquad 1 \geq \epsilon \sum_{k=1}^{n} (m+k)^{-1} \; P(T = \infty).$$

Since $\Sigma_k (m+k)^{-1}$ diverges to ∞, the probability that $T = \infty$ must be zero. This completes the proof of (a). The proof of (b) is similar. □

Corollary 3.1: Suppose X is the almost sure limit of an urn process corresponding to a continuous urn function f. Then $X = f(X)$ almost surely.

Proof: For each $\epsilon > 0$,

$$
\begin{aligned}
P(X \text{ in } A\epsilon) &= P(\{X_n\} \text{ eventually in } A\epsilon) \\
&= 0 .
\end{aligned}
$$

The first equality is due to the fact that X_n converges to X and $A\epsilon$ is open. The second follows from Proposition 3.1 and the Markov property. Similarly, $P(X \text{ in } B\epsilon) = 0$.

Thus, for each $\epsilon > 0$, $P(|X - f(X)| \leq \epsilon) = 1$, so $P(X = f(X)) = 1$. □

4. Downcrossings

Suppose the urn function f is continuous and $\{p_0\} = \{p: f(p) = p\}$, so p_0 must be a downcrossing. Special cases of processes with such urn functions include the Bernoulli process and the urn of Bernard Friedman. Theorem 2.1 and Corollary 3.1 together show that such urn processes must converge to p_0 almost surely. In this section, this result is generalized in two directions. Theorem 4.1 concerns a class of urn functions which are not necessarily continuous, but for which the associated urn processes converge to a single point almost surely. Theorem 4.2 refines Corollary 3.1 by showing that downcrossings must be in the support of the limit variable X, at least for urn functions which are continuous near the downcrossing and which map the open interval (0,1) into itself.

Theorem 4.1: Suppose f is an urn function, and there exist a point p_o in $(0,1)$ and a continuous urn function g satisfying:

i) $\{p_o\} = \{p: g(p) = p\}$

ii) for $p < p_o$, $f(p) \geq g(p)$ and
for $p > p_o$, $f(p) \leq g(p)$.

Then X_n converges to p_o almost surely.

Proof: Fix $\delta > 0$ and let g_δ be a continuous function satisfying:

a) $g_\delta = g$ on $[0, p_o - \delta]$

b) $g_\delta \leq g$ on $[p_o - \delta, p_o]$

c) $g_\delta = 0$ on $[p_o, 1]$

d) g_δ has a single downcrossing, say p_δ, which is in the interval $(p_o - \delta, p_o)$.

Then, $g_\delta \leq f$ on $[0,1]$. Now let $\{Y_n\}$ be an urn process with urn function g_δ and the same initial composition as $\{X_n\}$. As discussed above, Y_n converges to p_δ almost surely. By lemma 2.2 it may be assumed that $X_n \geq Y_n$ for all n, so X_n is eventually greater than $p_o - \delta$ almost surely. Similarly, it can be shown that X_n is eventually less than $p_o + \delta$ almost surely. Since δ is arbitrary, X_n converges to p_o almost surely. □

Example 4.1: A simple example of an urn function satisfying the conditions of Theorem 4.1 is the following step function:

$$f(p) = \begin{cases} 3/4 & \text{for } p < 1/2 \\ 1/4 & \text{for } p > 1/2 \end{cases}$$

The following localization lemma prepares the way for Theorem 4.2. To state the lemma, define y in $(0,1)$ to be <u>attainable at time k</u> if $P[X_k = y] > 0$. Suppose the urn function maps $(0,1)$ into itself. Then for an initial urn composition (x,m) with $0 < x < 1$, the set of states attainable at time k is just $\{(mx+j)(m+k)^{-1}: j = 0,\ldots,k\}$, and does not depend on the urn function.

<u>Lemma 4.1</u>: Suppose $\{X_n\}$ and $\{Y_n\}$ are urn processes with urn functions f and g respectively, and suppose f and g agree on a neighborhood N of the point p_o in $[0,1]$.

a) Suppose Y_n does not converge to p_o with positive probability for any initial urn composition (x,m). Then the same is true for X_n.

b) Suppose f and g map $(0,1)$ into itself, and $\{X_n\}$ and $\{Y_n\}$ have the same initial urn composition (x,m) with $0 < x < 1$. Then X_n converges to p_o with positive probability if and only if Y_n does.

<u>Proof:</u> Suppose $P_{(x,m)}[X_n$ converges to $p_o] > 0$. Then there exists k such that

(4.1) $P_{(x,m)}[X_n$ converges to p_o and X_ℓ in N for all $\ell \geq k] > 0$.

Since there are only a finite number of states attainable at time k, there must also exist j such that

(4.2) $P_{(x,m)}[X_n$ converges to p_o, X_ℓ in N for all $\ell \geq k$, and $X_k = j(m+k)^{-1}] > 0$.

Hence,

(4.3) $P_{(x,m)}[X_n$ converges to p_o, X_ℓ in N for all $\ell \geq k \mid X_k = j(m+k)^{-1}] > 0$

and

(4.4) $P_{(x,m)}[X_k = j(m+k)^{-1}] > 0$.

To prove a), note that (4.3) is equivalent to

(4.5) $P_{(j,m+k)}(X_n$ converges to p_o, X_ℓ in N for all $\ell \geq 0) > 0$.

Since f and g agree on N, (4.5) remains true when 'X' is replaced by 'Y', a contradiction to the assumption in a).

To prove b), note that inequality (4.4) just means that $j(m+k)^{-1}$ is attainable at time k for $\{X_n\}$, and hence for $\{Y_n\}$, since they share the same initial urn composition. Thus (4.4) holds when 'X' is replaced by 'Y'; (4.3) does also, since the urn functions f and g agree on N. Thus (4.2) holds with 'X' replaced by 'Y'. In particular, Y_n converges to p_o with positive probability. □

<u>Theorem 4.2</u>: Suppose f is an urn function which maps $(0,1)$ into itself, and f is continuous in a neighborhood of p_o, a downcrossing of f. Let $\{X_n\}$ be an urn process with urn function f and initial urn composition (x,m), $0 < x < 1$. Then X_n converges to p_o with positive probability.

<u>Proof</u>: Choose a neighborhood N of p_o in which f is continuous, and such that $\{p_o\} = N \cap C$. Construct a continuous urn function g such that g agrees with f in N, and $\{p_o\} = \{p: g(p) = p\}$. Then if $\{Y_n\}$ is an urn process with urn function g and initial urn composition (x,m), Y_n converges to p_o almost surely. By Lemma 4.1, then, X_n converges to p_o with positive probability. □

5. <u>Upcrossings</u>.

The following result is proved in this section.

<u>Theorem 5.1</u>: If p_o is an upcrossing point, then $P[X_n \to p_o] = 0$.

It is sufficient to prove the proposition when

$$(5.1) \qquad \begin{aligned} f(p) &\le p \quad \text{for } p < p_o, \\ &\ge p \quad \text{for } p > p_o. \end{aligned}$$

The reason is that, if p_o is an upcrossing point for an arbitrary g, then g agrees in a neighborhood of p_o with an f satisfying (5.1).

By Lemma 4.1 the proposition holds for g if and only if it holds for f. So (5.1) is assumed for the remainder of this section. However, it need not be assumed that f is continuous.

By (3.1) and (5.1) together with the Markov property, the urn process $\{X_n\}$ associated with f satisfies

$$(5.2) \qquad \begin{aligned} E[X_{n+1}|X_1,\ldots,X_n] &\le X_n \text{ a.s. on } [X_n < p_o], \\ &\ge X_n \text{ a.s. on } [X_n > p_o]. \end{aligned}$$

Any stochastic process $\{X_n\}$ satisfying (5.2) is said to be split by p_o.

Lemma 5.1: Let $\{X_n\}$ be a split process such that

(i) $\sup E|X_n| < \infty$,

(ii) $\lim (X_{n+1} - X_n) = 0$.

Then $\{X_n\}$ converges almost surely.

Proof: Suppose $\{X_n\}$ is split by p_o, and define

$$(5.3) \qquad Y_n = |X_n - p_o| .$$

Then $\sup_n E|Y_n| < \infty$ and

$$\begin{aligned} E[Y_{n+1}|X_1,\ldots,X_n] &= E[|X_{n+1} - p_o| \; |X_1,\ldots,X_n] \\ &\ge |E[X_{n+1}|X_1,\ldots,X_n] - p_o| \\ &\ge |X_n - p_o| \\ &= Y_n . \end{aligned}$$

The second inequality is a consequence of (5.2). It now follows from the submartingale convergence theorem that $\{Y_n\}$ converges almost surely. By (ii) and (5.3), so does $\{X_n\}$. □

Neither assumption (i) nor assumption (ii) can be deleted from Lemma 5.1. However, it seems likely that (ii) could be dropped if (i) were replaced by the stronger assumption of uniform boundedness or even just boundedness in L_2.

At this point, there is a temptation to develop further the theory of processes split by p_o in an attempt to give conditions for their nonconvergence to p_o and then to deduce Theorem 5.1 from them. We have successfully resisted this temptation. A second possible approach to a proof of Theorem 5.1 would be to make a pathwise comparison of $\{X_n\}$ with a Polya process $\{Y_n\}$. If $p_o = 1/2$, techniques like those used to prove Lemma 2.2 lead to the realization of $\{X_n\}$ and $\{Y_n\}$ on a common probability space so as to satisfy

$$|X_n - 1/2| \geq |Y_n - 1/2|$$

for all n. Since $\{Y_n\}$ converges to 1/2 with probability zero, the same is true of $\{X_n\}$. Unfortunately, this simple argument does not obviously generalize when $p_o \neq 1/2$. Instead of pursuing the idea further, we will use the gambling theoretic methods of Dubins and Savage (1965) to make a quite different sort of comparison of the two processes. Some of the notation and terminology below is taken from Dubins and Savage.

Consider a gambling problem in which the fortunes are pairs (x,m) where $0 < x < 1$ and m is a nonnegative integer. To describe the gambles available, first let $0 \leq q \leq 1$ and define q to be <u>compatible</u> with x if

$$0 \leq q \leq x \leq p_o \quad \text{or} \quad p_o \leq x \leq q \leq 1. \tag{5.4}$$

To each triple (q,x,m) is associated the gamble $\gamma = \gamma(q,x,m)$ which is the distribution of

$$((mx + Y)(m+1)^{-1},\ m + 1)$$

where $P[Y=1] = q = 1 - P[Y=0]$. The gambles available at (x,m) are all $\gamma(q,x,m)$ where q is compatible with x. In effect, the gambler is permitted, at each stage of play, to select a q compatible with the proportion of red balls in the urn and then to add a red ball with probability q and a black with probability $\bar{q} = 1-q$.

A strategy σ at (x,m) can here be regarded as a sequence $q_o, q_1, \ldots$ where q_o is a constant compatible with x and, for every finite sequence $x_1,\ldots,x_n$ of elements of $(0,1)$, $q_n(x_1,\ldots x_n)$ is compatible with x_n. The strategy determines the distribution of a stochastic process $(X_1,\ m+1)$, $(X_2,\ m+2),\ldots$ by specifying that the distribution of $(X_1,\ m+1)$ is $\gamma(q_o,\ x,\ m)$ and that the conditional distribution of $(X_{n+1},\ m+n+1)$ is $\gamma(q_n(x_1,\ldots,x_n),x_n,\ m+n)$ given $X_1 = x_1,\ldots,X_n = x_n$. Since the sequence of second coordinates is deterministic, only the X_n 's will be mentioned in the sequel.

Every function f satisfying (5.1) determines a strategy at (x,m) which has $q_o = f(x)$ and $q_n(x_1,\ldots,x_n) = f(x_n)$ for all $x_1,\ldots,x_n$. Under this strategy the process $\{X_n\}$ obviously has the same distribution as the urn process associated with f. In particular, the <u>Polya strategy</u> at (x,m) is determined by the identity function and, under the Polya strategy, $\{X_n\}$ is a Polya process.

The utility function for the gambling problem is defined to be

$$u(x,m) = \varphi(x)$$

where

$$\varphi(x) = \varphi_{\alpha+1,\beta+1}(x) = c_{\alpha+1,\beta+1}x^{\alpha}(1-x)^{\beta}. \tag{5.5}$$

The function φ is a Beta density with parameters $\alpha + 1$, $\beta + 1$.

Notice that, under every available strategy σ, the process $\{X_n\}$ satisfies all the hypothesis of Lemma 5.1 and, therefore, converges almost surely to a limiting random variable X. The utility of σ is set equal to the expected utility of the limit as below.

$$
\begin{aligned}
(5.6) \qquad u(\sigma) &= \int \lim_n u(X_n, m+n)\, d\sigma \\
&= \int \varphi(X) d\sigma .
\end{aligned}
$$

This definition coincides with that of Dubins and Savage (1965, Formula 3.2.1) as follows from Theorem 3.2 of Sudderth (1971).

After a simple lemma on Beta distributions, it will be shown that the Polya strategy is optimal whenever φ has mode p_o.

<u>Lemma 5.2</u>. Let $\{\alpha_n\}$ and $\{\beta_n\}$ be sequences of positive numbers such $\alpha_n + \beta_n \to \infty$ while $\alpha_n(\alpha_n + \beta_n)^{-1} \to p$. Then the corresponding sequence of Beta $(\alpha_n + 1, \beta_n + 1)$-distributions converges in distribution to a point mass at p_o. Furthermore, $\sup \varphi_{\alpha_n + 1,\, \beta_n + 1} \to \infty$.

<u>Proof</u>: The mean of a Beta $(\alpha_n + 1, \beta_n + 1)$ is $(\alpha_n + 1)(\alpha_n + \beta_n + 2)^{-1}$, which converges to p. The variance is $(\alpha_n + 1)(\beta_n + 1)(\alpha_n + \beta_n + 2)^{-2}(\alpha_n + \beta_n + 1)^{-1}$, which converges to zero. This proves the first assertion.

Because their variances are converging to zero, the densities cannot remain bounded, and the second assertion follows. □

<u>Lemma 5.3</u>. If the mode $\alpha(\alpha + \beta)^{-1}$ of φ is equal to p_o, then the Polya strategy at (x,m) is optimal in the sense that it achieves the supremum

$$W(x,m) = \sup u(\sigma)$$

taken over all strategies σ at (x,m).

Proof: Let $Q(x,m)$ be the utility of the Polya strategy at (x,m). Then, by (5.6) and Example 1.3,

$$(5.7)\qquad Q(x,m) = \int_0^1 \varphi_{\alpha+1,\beta+1}(t)\varphi_{s,m-s}(t)dt = c\int_0^1 \varphi_{\alpha+s,\ \beta+m-s}(t)dt = c\ ,$$

where $c = (c_{\alpha+1,\beta+1}\ c_{s,m-s})(c_{\alpha+s,\ \beta+m-s})^{-1}$.

It suffices to show $Q \geq W$. This inequality will follow from Dubins and Savage (1965)(the fourth paragraph of section 3.3) or from Lemma 4.8 of Dubins and Sudderth (1977) once it is verified that

$$(5.8)\qquad \gamma(q,x,m)Q \leq Q(x,m)$$

for all (x,m) and all q compatible with x, and that

$$(5.9)\qquad Q(\sigma) \geq u(\sigma)$$

for all available strategies σ, where

$$(5.10)\qquad Q(\sigma) = \int \lim_n Q(X_n,\ m+n)d\sigma.$$

To prove (5.8), first calculate its left-hand-side as follows:

$$(5.11)\qquad \begin{aligned}\gamma(q,x,m)Q &= qQ((s+1)(m+1)^{-1},m+1) + \bar{q}Q(s(m+1)^{-1},\ m+1)\\ &= \int_0^1 \varphi_{\alpha+1,\beta+1}(t)\{q\varphi_{s+1,m-s}(t) + \bar{q}\varphi_{s,m-s+1}(t)\}dt\\ &= c\int_0^1 \varphi_{\alpha+s,\beta+m-s}(t)\{qmts^{-1} + \bar{q}m\bar{t}(m-s)^{-1}\}dt\\ &= mc\{q(\alpha+s)s^{-1}(\alpha+\beta+m)^{-1} + \bar{q}(\beta+m-s)(m-s)^{-1}(\alpha+\beta+m)^{-1}\},\end{aligned}$$

where the first equality is by definition of $\gamma(q,x,m)$, the second is by (5.7), the third uses the definition of a Beta density, and the final equality uses the formula for the mean of a Beta density.

Let $\psi(q) = \gamma(q,x,m)Q - Q(x,m)$. By (5.7) and (5.11), ψ is a linear function of q with slope

$$\psi'(q) = mc(\alpha+\beta+m)^{-1}\{(\alpha + s)s^{-1} - (\beta+m-s)(m-s)^{-1}\}.$$

Thus

$$\begin{aligned}\psi'(q) > 0 &\Leftrightarrow (\alpha+s)s^{-1} > (\beta+m-s)(m-s)^{-1} \\ &\Leftrightarrow \alpha\beta^{-1} > s(m-s)^{-1} \\ &\Leftrightarrow \alpha(\alpha+\beta)^{-1} > sm^{-1}.\end{aligned}$$

Also, it is easy to check that

$$\psi(x) = 0.$$

By assumption, one of the two chains of inequalities in (5.4) holds. If it is the first, then $q \le x = sm^{-1} < p_o = \alpha(\alpha+\beta)^{-1}$. So $\psi'(q) > 0$. Hence, $\psi(q) < \psi(x) = 0$. If, on the other hand, $p_o < x \le q$, then $\psi'(q) < 0$ and again $\psi(q) < \psi(x)$. Therefore, in all cases, $\psi(q) \le 0$ which completes the proof of (5.8).

To prove (5.9), let σ be an available strategy. Then, under σ, X_n converges to X almost surely and, by Lemma 5.2, the Beta $((m+n)X_n,(m+n)\bar{X}_n)$-distributions converge almost surely to point mass at X. Let Qx_n be the density for a Beta $((m+n)X_n, (m+n)\bar{X}_n)$. Then

$$
\begin{aligned}
Q(\sigma) &= \int \lim_n Q(X_n, m+n)\,d\sigma \\
&= \int \{\lim_n \int_0^1 \varphi(t)\varphi_{X_n}(t)\,dt\}\,d\sigma \\
&= \int \varphi(X)\,d\sigma \\
&= u(\sigma).
\end{aligned}
$$

This completes the proof of Lemma 5.3. □

Turn now to the proof of Theorem 5.1. Suppose, by way of contradiction, that

$$P[X_n \to p_o] = \epsilon > 0.$$

Let σ be the strategy associated with f. Then, by (5.6),

$$u(\sigma) \geq \epsilon\ \varphi_{\alpha+1,\beta+1}(p_o).$$

Let $\alpha + \beta \to \infty$ while $\alpha(\alpha+\beta)^{-1} = p_o$. It follows from the final assertion of Lemma 5.2 that $u(\sigma)$ approaches infinity. By (5.7) and the first assertion of Lemma 5.2,

$$
\begin{aligned}
Q(x,m) &= \int_0^1 \varphi_{\alpha+1,\beta+1}\varphi_{s,m-s} \\
&\to \varphi_{s,m-s}(p_o).
\end{aligned}
$$

However, by Lemma 5.3,

$$u(\sigma) \leq Q(x,m),$$

which gives the desired contradiction.

6. Extensions and questions. For which urn functions do the associated urn processes converge? For convergent urn processes, what is the support of the limit variable X? We cannot yet give complete answers to these questions.

Theorem 6.1 provides the answer to both questions for a particular class of continuous urn function. As before, let $C = \{p: f(p) = p\}$ and let U and D be the set of upcrossings of f and downcrossings of f respectively.

Theorem 6.1: Suppose f is a continuous urn function which maps $(0,1)$ into itself, and $C = U \cup D$. Then the urn process $\{X_n\}$ with urn function f and initial urn composition (x,m), $0 < x < 1$, converges almost surely to a random variable X, and the support of X is D.

Proof: By Theorem 2.1 X_n converges to X, whose support is contained in C by Corollary 3.1. By Theorem 4.2, each point of D is in this support, while by Theorem 5.1 no point of U is. $\square$

The assumption of continuity made in Theorem 6.1 can, with some effort, be relaxed to piecewise continuity or even further. Also, if the crossing set C contains a nondegenerate interval, it can be shown that the support of the limit variable X must contain the interval as well.

Here is a puzzle which we have not been able to solve. Call a point $p_0 \in C$ a touchpoint if, for some $\varepsilon > 0$ other $f(p) > p$ for all p such that $0 < |p-p_0| < \varepsilon$, or $f(p) < p$ for all such p. Do touchpoints belong to the support of X?

References

1. Dubins, Lester E. and Savage, Leonard J. (1965), *How to Gamble if You Must: Inequalities for Stochastic Processes* (Reissued by Dover, New York, 1976).

2. Dubins, Lester E. and Sudderth, William D. (1977), "Countably Additive Gambling and Optimal Stopping," *Zeitschrift fur Wahrscheinlichkeitstheorie* 41, 59-72.

3. Freedman, David A. (1965), "Bernard Friedman's Urn," *Annals of Mathematical Statistics*, 36, 956-970.

4. Johnson, Norman L. and Kotz, Samuel (1977). *Urn Models and Their Application* (Wiley, New York).

5. Sudderth, William D. (1971), "On Measurable Gambling Problems," *Annals of Mathematical Statistics*, 42, 260-269.